

СЕМЕЙСТВО МНОГОПРОЦЕССОРНЫХ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ С
ДИНАМИЧЕСКИ ПЕРЕСТРАИВАЕМОЙ АРХИТЕКТУРОЙ

Н.Н. Дмитренко¹, И.А. Каляев¹, И.И. Левин¹, Е.А. Семерников²

НИИ многопроцессорных вычислительных систем имени академика А.В. Каляева

Южного федерального университета, г. Таганрог, Россия¹

Южный научный центр Российской академии наук, г. Ростов-на-Дону, Россия²

Аннотация

В статье приводятся принципы построения и описание семейства многопроцессорных вычислительных систем с динамически перестраиваемой архитектурой на основе реконфигурируемой элементной базы производительностью до 6 Тфлопс, создаваемого в рамках Федеральной целевой программы «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007-2012 годы». В качестве основного вычислительного элемента в них используются не универсальные микропроцессоры, а программируемые логические интегральные схемы сверхбольшой интеграции. Все представители семейства предназначены для решения вычислительно трудоемких задач различных предметных областей и обеспечивают на всех задачах реальную производительность не ниже 50% от пиковой производительности.

СУПЕРЭВМ, РЕКОНФИГУРИРУЕМЫЕ ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ, ДИНАМИЧЕСКИ ПЕРЕСТРАИВАЕМАЯ АРХИТЕКТУРА, ПЛИС, ПАРАЛЛЕЛЬНОЕ ПРОГРАММИРОВАНИЕ.

Abstract

In the article are given principles of design and description of multiprocessor computer systems with dynamically reconfigurable architecture on basis of reconfigurable elementary base with performance up to 6 TFlops, which is created according to Federal task program “Research and design of priority directions of Russian scientific and technological complex development during 2007-2012”. As basic computing elements there are used not general purpose microprocessors, but programmable logical integrated super large scale integration circuits. All representatives of the family are supposed to be used for solving tasks of high computational complexity from various subject fields and provide real performance not less than 50% from the peak one for all tasks.

SUPERCOMPUTER, RECONFIGURABLE COMPUTER SYSTEMS, DYNAMICALLY RECONFIGURABLE ARCHITECTURE, FPGA, PARALLEL PROGRAMMING.

Введение

Наиболее распространенные в мире кластерные МВС имеют высокие значения пиковой производительности. В то же время пользователи отмечают принципиальные недостатки кластерных МВС, которые связаны с относительно низкой скоростью процедур межпроцессорного обмена, ограниченной пропускной способностью сети передачи данных, необходимостью синхронизации множества взаимосвязанных последовательных процессов, каждый из которых выполняется на отдельном процессоре, и т.д. Все это приводит к тому, что высокую реальную производительность кластерные суперЭВМ демонстрируют, в основном, только при решении класса слабосвязанных задач, не требующих большого числа информационных обменов, в то время как при решении задач других классов их реальная производительность существенно снижается и не превышает 5-10% от декларируемой пиковой производительности системы. Это является следствием неадекватности данной конкретной архитектуры суперкомпьютера информационной структуре решаемой задачи [1, 2, 3, 4].

Недостатки кластерных систем, связанные с их неизменной «жесткой» архитектурой, позволяет преодолеть концепция создания МВС с «гибкой», динамически перестраиваемой (программируемой) архитектурой на основе реконфигурируемой элементной базы или иначе – реконфигурируемые вычислительные системы (РВС) [3, 4]. Концепция создания РВС разработана и развивается в Научно-исследовательском институте многопроцессорных вычислительных систем имени академика А.В. Каляева Южного федерального университета (НИИ МВС ЮФУ) уже более 20 лет. В отличие от многопроцессорных вычислительных систем с «жесткой» архитектурой архитектура РВС может динамически изменяться в процессе их функционирования. В результате у пользователя появляется возможность адаптации архитектуры вычислительной системы под структуру решаемой им задачи. Иными словами, пользователь может, оставаясь в рамках базовой архитектуры системы, создавать проблемно-ориентированные вычислители, структура которых соответствует структуре решаемой задачи. Проведенные исследования и многочисленные практические разработки, выполненные в НИИ МВС ЮФУ, показали [5,6,7], что реализация данной концепции обеспечивает высокую реальную производительность РВС, близкую к пиковой на широком классе задач, в том числе при решении «сильносвязанных» задач. В качестве основного вычислительного элемента в РВС используются не универсальные микропроцессоры, а программируемые логические интегральные схемы (ПЛИС) сверхбольшой интеграции. На основе предложенной концепции возможно без привлечения больших финансовых затрат создавать высокопроизводительные вычислительные системы, существенно опережающие зарубежные суперЭВМ по таким характеристикам как соотношения «реальная производительность/пиковая производительность» (эффективность) и «реальная производительность/объем» (компактность).

Еще одной важной характеристикой РВС является практически линейный рост производительности при наращивании аппаратного ресурса [4], что является следствием исключения из вычислительного процесса множества вспомогательных действий, которые неизбежно сопровождают вычисления в универсальных процессорах, на базе которых построены суперЭВМ традиционной архитектуры.

Архитектурные особенности и принципы организации вычислений в РВС

В отличие от традиционных методов организации параллельных вычислений РВС ориентируются на абсолютно параллельную форму алгоритма задачи – ее информационный граф. Под информационным графом [2] понимается граф, вершины которого соответствуют арифметико-логическим операциям над операндами или ячейками (каналами) памяти, в которых расположены информационные массивы. Дуги информационного графа соответствуют информационной зависимости между вершинами. Если существует направленная дуга, соединяющая вершину a с вершиной b , то это означает, что результат операции, соответствующий вершине информационного графа a , является входным операндом операции, соответствующей вершине b . Вершины, соответствующие каналам памяти, в которых располагается входная и выходная информация, являются входными и выходными вершинами информационного графа. Информационные графы больших и сложных задач могут содержать миллионы вершин.

Идея концепции построения РВС заключается в аппаратной реализации всех операций, предписанных вершинами информационного графа задачи, всех каналов передачи данных между вершинами, соответствующих дугам графа, и всех информационных каналов, соответствующих входным и выходным вершинам. В этом случае задача, определенная информационным графом, будет выполнена максимально быстро, поскольку обеспечивается максимально возможное распараллеливание вычислений. Такое решение задачи принято называть *структурным* [4].

Основными вычислительными блоками в РВС являются *макропроцессоры* (МАП). Макропроцессор позволяет реализовывать крупные операции, которые предписаны вершинами информационного графа. Макропроцессор представляет собой некоторый набор элементарных процессоров (ЭП), объединяемых в единый программно-неделимый вычислительный ресурс с помощью локального пространственного коммутатора (K_1) (рис. 1, *a*). Внутренний коммутатор макропроцессора, как правило, соединяет по полному графу все информационные входы и выходы элементарных процессоров, внешние информационные входы X , внешние информационные выходы Z макропроцессора. Настройка макропроцессора на крупную функционально законченную операцию (макрооперацию) производится с помощью блока

макроопераций (БМ). В каждый момент времени макропроцессор может реализовывать только одну макрооперацию. Недействующие в макрооперации элементарные процессоры будут простаивать.

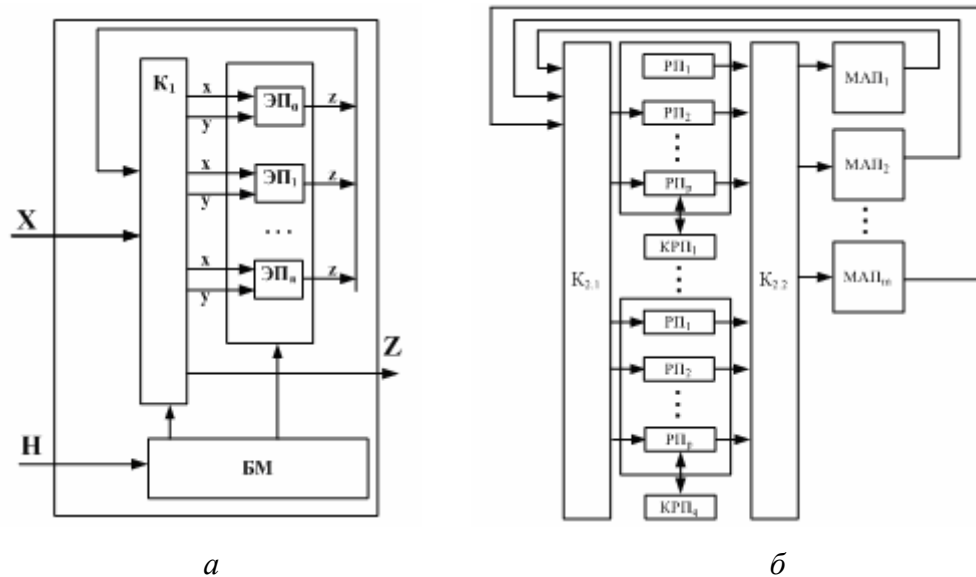


Рис.1. Структурная схема макропроцессора (а) и обобщенная схема РВС (б)

Следует отметить, что в отличие от стандартного микропроцессора элементарный процессор макропроцессора не управляет процессом обработки информации, а лишь реализует соответствующую ему операцию над операндами, поступающими на его входы.

Для реализации того или иного информационного графа решаемой задачи макропроцессоры должны иметь возможности соединения в вычислительные параллельно-конвейерные структуры. С этой целью в состав РВС включен *коммутатор второго уровня* или *системный коммутатор*, обеспечивающий различные варианты соединения макропроцессоров друг с другом. Кроме того, в состав РВС входит распределенная память, обеспечивающая возможности параллельной выдачи массива входных данных на входы макропроцессоров и записи результатов вычислений с их выходов.

Обобщенная схема РВС приведена на рис.1 б. В ее состав входят: множество макропроцессоров, состоящих из элементарных процессоров, соединяемых между собой с помощью локального коммутатора; распределенная память (РП), включающая в свой состав контроллер распределенной памяти (КРП) и множество каналов РП, а также системный коммутатор $К_2$, предназначенный как для соединения макропроцессоров с друг с другом в вычислительные структуры, так и для их параллельного доступа к каналам распределенной памяти. Множество входных и выходных дуг информационного графа реализуется системным коммутатором $К_2$, обеспечивающим подключение каналов распределенной памяти к тем или иным входам и выходам макропроцессоров. При этом массивы входных данных отображаются

в соответствующие каналы распределенной памяти, процедура обращения к которым реализуется с помощью КРП.

Основные положения концепции были сформулированы академиком А.В. Каляевым в середине 80-х годов прошлого века и развиваются в НИИ МВС ЮФУ уже более двадцати лет. Опираясь на сформулированную концепцию, в 80-90-е годы в НИИ МВС был создан целый ряд вычислительных систем с реконфигурируемой архитектурой. Созданные системы представляли собой первые шаги на пути практического становления данного направления вычислительной техники.

Несмотря на большой объем проведенных 80–90-е годы исследований в области высокопроизводительных вычислительных систем с реконфигурируемой архитектурой и создание ряда опытных образцов, дело до серийного выпуска систем такого типа не дошло. Это было вызвано тем, что реконфигурируемые вычислительные системы, созданные на основе элементной базы того времени, не позволяли в полной мере реализовать преимущества концепции. Перестройка архитектуры затрагивала только связи между процессорами, в то время как структура самих процессоров перестраивалась в узких пределах и была достаточно жесткой, поскольку процессоры были выполнены на серийных микропроцессорах с неизменной структурой. Вычислительные системы получались громоздкими, малонадежными, сложными в эксплуатации.

Таким образом, реализация и широкое внедрение в практику идеи высокопроизводительных вычислительных систем с реконфигурируемой архитектурой сдерживалась отсутствием в нашей стране соответствующей этой концепции элементной базы. Полномасштабная же реализация РВС требует наличия принципиально иной элементной базы, удовлетворяющей следующим основным критериям:

- высокой степени интеграции, позволяющей реализовать крупные фрагменты вычислений;
- сравнительно низкой стоимости при мелкосерийном производстве вычислительных систем;
- возможности аппаратной реализации специализированных вычислительных структур;
- возможности реализации различных схем распараллеливания вычислений;
- возможности реконфигурации сформированных вычислительных структур при переходе от одних алгоритмов и схем распараллеливания к другим;
- поддержки разработки схемотехнических решений системами автоматизированного проектирования.

Всем этим требованиям отвечают появившиеся в конце XX-го и начале XXI-го века программируемые логические интегральные схемы (ПЛИС) со сверхвысокой степенью

интеграции (в английской аббревиатуре FPGA – Filed Programmable Gates Array). В ПЛИС изначально заложены возможности реконфигурирования их внутренней архитектуры, и поэтому они наилучшим образом соответствуют концепции реконфигурируемых вычислительных систем.

Основными преимуществами программируемых логических интегральных схем являются: возможность аппаратной реализации сложных параллельных алгоритмов; наличие средств САПР, позволяющих провести полное моделирование системы; возможность программирования или изменения конфигурации непосредственно в системе; совместимость различных проектов при их переводе на языки описания аппаратуры - VHDL, AHDL, Verilog и др.

ПЛИС выпускаются несколькими фирмами: Xilinx, Altera, Atmel, Vantis, Lucent и др. Ведущим производителем ПЛИС является фирма Xilinx, занимающая более 40% мирового рынка. При этом фирма каждые 1,5-2 года выпускает на рынок все новые поколения ПЛИС, превосходящие показатели предыдущих поколений по потребляемой мощности, тактовой частоте и функциональным возможностям (рис. 2).

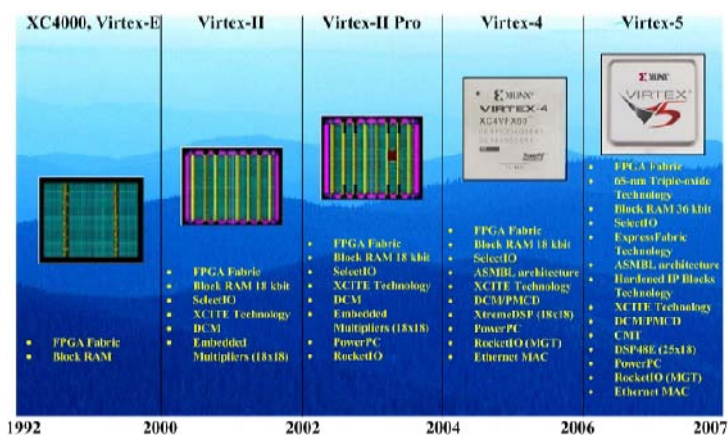


Рис. 2. Расширение функциональных возможностей высокопроизводительных ПЛИС фирмы Xilinx

На рис.3 показаны графики, отражающие основные тенденции развития высокопроизводительных ПЛИС фирмы Xilinx.

Как видно, стоимость ПЛИС имеет тенденцию к снижению в абсолютном выражении. Производительность (синий график) оценивалась путем размещения в ПЛИС максимально возможного количества 32-разрядных устройств с плавающей запятой стандарта IEEE-754, функционирующих на частоте работы микросхемы. Красный график показывает экспоненциальный рост отношения «производительность/стоимость», что, в свою очередь, свидетельствует о росте экономической эффективности элементной базы на основе ПЛИС.

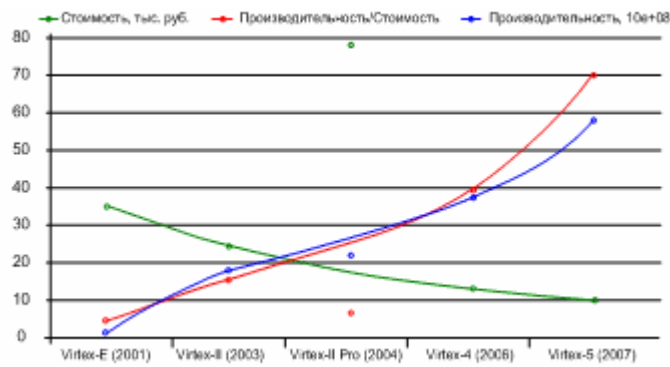


Рис. 3. Основные тенденции развития высокопроизводительных ПЛИС фирмы Xilinx

Таким образом, наличие широких возможностей реконфигурирования, поддержанных удобными САПР, и высокая экономическая эффективность современных ПЛИС позволяют сделать однозначный вывод о том, что они являются наиболее перспективной элементной базой для построения РВС.

Применение ПЛИС большой степени интеграции дало возможность перейти к созданию больших решающих полей, объединяющих вычислительные ресурсы множества ПЛИС в единый вычислительный ресурс. Это позволило отказаться от прежнего подхода, в котором отдельным микросхемам отводилось определенное постоянное функциональное назначение, и перейти к более прогрессивной архитектуре, когда в любой микросхеме может быть реализовано любое функциональное устройство, которое необходимо в данном месте решающего поля. В этом смысле решающее поле является как бы аналогом сверхбольшой ПЛИС, объединяющей в себе все их аппаратные ресурсы.

Однако построение больших решающих полей на ПЛИС выявило новые проблемы. Первая проблема - это негативный эффект границ, возникающих на стыках отдельных ПЛИС при их объединении в решающее поле. Эффекты границ сказываются при реализации больших фрагментов вычислительных структур, которые не могут быть реализованы в пределах отдельной микросхемы, а требуют ресурса нескольких ПЛИС. Вторая проблема - это конструктивно-технологические ограничения, которые заключается в том, что невозможно разместить неограниченное количество ПЛИС на печатной плате приемлемого размера. Эта проблема решается путем модульного построения аппаратных средств РВС и модульной наращиваемости на основе унифицированных базовых модулей. Базовые модули не только позволяют путем комплексирования создавать РВС необходимой производительности, но и сами по себе являются реконфигурируемой вычислительной системой, способной совместно с персональным компьютером (в качестве интеллектуального терминала) решать пользовательские задачи.

Однако введение модульного принципа построения аппаратных средств обострило первую проблему, породив еще один тип границ – межмодульный. Суть решения проблемы границ заключается в создании единого типа интерфейса как между ПЛИС в пределах базового модуля, так и между ПЛИС разных базовых модулей.

Ограниченность аппаратного ресурса приводит к тому, что для больших задач весь информационный граф не может быть отображен в имеющемся аппаратном ресурсе РВС. В этом случае информационный граф сегментируется на фрагменты – непересекающиеся подграфы, физически реализуемые в аппаратуре РВС, а решение большой задачи выполняется *структурно-процедурным* способом, при котором на решающее поле ПЛИС поочередно отображаются подграфы информационного графа, и вычисления в соответствии с отображенным подграфом выполняются структурно, а смена подграфов выполняется процедурно [4, 7, 8].

В простейшем случае структурно-процедурный вычислительный процесс в РВС организуется следующим образом. В аппаратном ресурсе РВС реализуется один из подграфов сегментированного информационного графа задачи и на его входы из блоков памяти подаются входные данные этого фрагмента. Процесс вычисления реализуется вычислительной структурой, соответствующей данному подграфу, и на выходе формируется результат вычислений, который запоминается в соответствующих блоках памяти. Далее в аппаратном ресурсе РВС реализуется следующий подграф сегментированного информационного графа задачи, и процесс повторяется. Этот процесс показан на рис. 4.

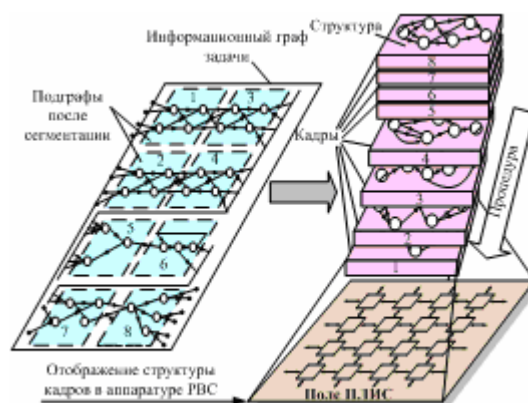


Рис. 4. Процесс решения задачи в РВС

При реализации структурно-процедурного метода информационный граф задачи представляется в виде кортежа изоморфных подграфов, которые являются информационно-независимыми или непосредственно зависящими друг от друга. Кортеж информационных подграфов преобразуется в специальную вычислительную конструкцию - *кадр*. Можно сказать, что кадру соответствует подграф задачи, реализованный аппаратно, через который следует поток операндов. При этом каждая группа операндов (результатов) соответствует входным

(выходным) вершинам определенного подграфа кортежа. Смена кадров в РВС осуществляется процедурно по единой для всей системы программе.

Механизм последовательного обхода подграфов информационного графа задачи кадрами принято называть *структурно-процедурной организацией вычислений*.

РВС со структурно-процедурной организацией вычислений является гибридом фон-неймановской архитектуры и архитектуры потока данных. Такая организация вычислений обеспечивает детерминизм выполнения программы, что в общем случае недостижимо в многопроцессорных системах, построенных по традиционной мультипроцедурной архитектуре (кластерные МВС). При этом обеспечивается также и высокая эффективность параллельных вычислений на широком классе задач.

Особенность архитектуры РВС заключается в том, что в процессе конструирования она не формируется окончательно, а остается в определенном смысле незавершенной и открытой. Окончательное программирование архитектуры РВС включает создание функциональных узлов (элементарных процессоров) для выполнения вычислений, настройку прямых информационных каналов в коммутаторах K_1 и K_2 между элементарными процессорами, настройку блоков распределенной памяти на реализацию процедур чтения и записи информационных массивов. Аппаратно-программные средства РВС позволяют синтезировать произвольные соединения между компонентами системы, создавая необходимые вычислительные структуры. Совокупность вычислительных структур, созданных в рамках базовой архитектуры РВС, образуют виртуальный проблемно-ориентированный вычислитель, структура которого адекватна информационному графу (подграфу) решаемой задачи [4].

На основе описанной выше концепции в НИИ МВС ЮФУ за последние 5-6 лет был создан ряд РВС различной производительности и назначения, из которых две системы в настоящее время выпускаются серийно. Показатели «компактности» и «эффективности» созданных систем составили 5-6 Гфлопс/дм³ и менее 18000-20000 руб/Гфлопс. Для задач символьной обработки эти же показатели составили 115×10^9 операций/дм³ и 2000 руб./млрд.оп.

Состав и характеристики семейства РВС

В настоящее время по заданию Федерального агентства по науке и инновациям ведется выполнение Государственного контракта № 02.524.12.4002 по теме «Создание семейства высокопроизводительных многопроцессорных вычислительных систем с динамически перестраиваемой архитектурой на основе реконфигурируемой элементной базы и их математического обеспечения для решения вычислительно трудоемких задач» в рамках Федеральной целевой программы «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007-2012 годы».

Целью разработки является создание на единых архитектурных принципах семейства программно-совместимых реконфигурируемых высокопроизводительных вычислительных систем производительностью от 0,025 Тфлопс до 6 Тфлопс.

В рамках государственного контракта будет создано семейство реконфигурируемых вычислительных систем:

- РВС-5 – высокопроизводительная система производительностью 6 Тфлопс;
- РВС-1Р и РВС-1К – системы производительностью более 1 Тфлопс;
- РВС-0.2-РС – рабочая станция производительностью 300 Гфлопс;
- РУПК-50 и РУПК-25 – ускорители персональных компьютеров производительностью 50 и 25 Гфлопс.

Старшие представители семейства создаются на принципах модульной наращиваемости и будут обладать почти линейным ростом реальной производительности в зависимости от увеличения аппаратного ресурса [9]. Состав семейства РВС показан на рис. 5.

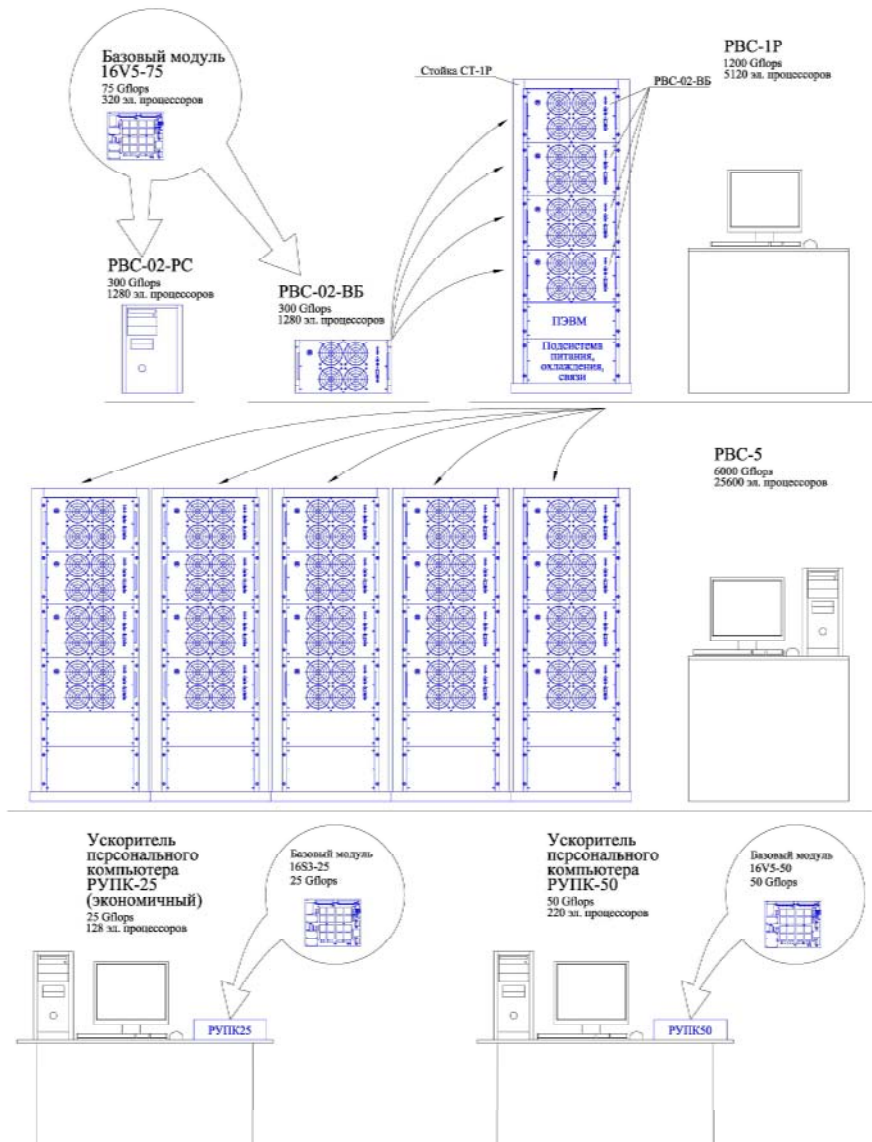


Рис. 5. Состав семейства РВС

Для создания семейства РВС, системного и прикладного программного обеспечения его представителей, информационной и программной инфраструктуры для обучения и поддержки потенциальных пользователей семейства образован консорциум из ряда научных и научно-производственных организаций, в который вошли: НИИ МВС ЮФУ (г. Таганрог) – головной исполнитель проекта; ФГУП «НИИ «Квант» (г. Москва); «Южный научный центр РАН» (г. Ростов-на-Дону); «Специальная астрофизическая обсерватория РАН» (пос. Нижний Архыз); НИЦ «СуперЭВМ и нейрокомпьютеров» (г. Таганрог); НИВЦ МГУ (г. Москва).

Реальная производительность всех представителей семейства РВС планируется не ниже 50% от указанной пиковой производительности на широком классе вычислительных задач. При этом они будут обладать существенно меньшими габаритами, энергопотреблением и стоимостью по сравнению с суперЭВМ кластерного типа аналогичной производительности.

Системы с высокой реальной производительностью позволят решить важнейшие научно-технические задачи и обеспечить прорывы в таких наукоемких технологических областях, как молекулярная фармакология, нанoeлектроника, создание энергетических комплексов нового поколения, а также проведение фундаментальных научных исследований в астрофизике, микробиологии, физике твердого тела, нейроматематике, томографическом исследовании приповерхностных слоев Земли и т.д.

Базовые модули семейства РВС

Базовые модули, из которых строятся РВС, в полной мере сохраняют все концептуальные особенности архитектуры реконфигурируемых систем. Семейство РВС строится всего из трех типов базовых модулей: 16V5-75, 16V5-50 и 16S3-25. Базовый модуль 16V5-75, как самый производительный, используется в моделях РВС-5, РВС-1Р и РВС-0.2-РС. Базовые модули 16V5-50 и 16S3-25 входят в состав ускорителей персональных компьютеров РУПК-50 и РУПК-25. Общие принципы построения базовых модулей одинаковы и поэтому рассмотрим более подробно базовый модуль 16V5-75.

Базовый модуль 16V5-75 обладает характеристиками, представленными в таблице.

Таблица – Характеристики базового модуля 16V5-75

Параметр	Значение
Производительность (64 разряда), Гфлопс	75
Производительность (32 разряда), Гфлопс	140
Потребляемая мощность, ВА	200
Объем оперативной распределенной памяти, Гбайт	1,25
ПЛИС решающего поля Virtex 5 XC5VLX110, шт.	16
Количество эквивалентных вентилях в ПЛИС, шт.	$8 \cdot 10^6$
Тактовая частота, МГц	250
Количество внешних LVDS каналов, шт.	224
Скорость межмодульного обмена, Мбит/сек	22000

Решающее поле базового модуля 16V5-75 выполнено на 16-ти ПЛИС Virtex 5 XC5VLX110-2FF1153 фирмы Xilinx, каждая из которых содержит около $8 \cdot 10^6$ эквивалентных вентилях.

Микросхемы расположены в узлах двумерной решетки 4 x 4 и соединены между собой ортогональной системой связей по близкодействию. Такая система связей позволяет существенно упростить печатную плату и улучшить ее частотные характеристики, поскольку связи между соседними ПЛИС не превышают четыре сантиметра. Данные между несмежными микросхемами передаются по транзитным каналам через промежуточные ПЛИС, используя систему ортогональных связей.

Связи между ПЛИС решающего поля базового модуля реализованы на основе стандарта LVDS. Преимуществами стандарта LVDS являются: низкая потребляемая мощность выходных каскадов, низкий уровень создаваемых электромагнитных излучений, невосприимчивость к синфазным электромагнитным помехам и наличие в микросхемах семейства Vertex 5 аппаратной поддержки для организации высокоскоростных передач данных на основе стандарта LVDS. Физически шины связи представляют собой набор пар дифференциальных полосковых передающих линий, с обоих концов подключенных к определенным выводам микросхем. Для надежной передачи данных по LVDS в 16V5-75 задействованы специальные ресурсы семейства микросхем Vertex 5, поддерживающие алгоритм оптимальной битной и кадровой синхронизации и позволяющие учесть все нюансы линии передачи.

Базовый модуль имеет подсистему обмена информацией по быстрым LVDS-каналам с другими базовыми модулями, которая предназначена для передачи промежуточных результатов вычислений непосредственно из микросхем данного модуля непосредственно в микросхемы решающего поля других базовых модулей. Передача осуществляется посредством специальных кабелей, подключаемых к соединителям типа QTE-056 (рис.6).

Единый тип интерфейса между ПЛИС как в пределах одного базового модуля, так и между различными базовыми модулями, обеспечивает одинаковые скорости обмена и снижает эффект границ в пределах общего схемотехнического ресурса PBC.

К периферийным ПЛИС решающего поля, расположенных на краях двумерной решетки 4x4, подключены микросхемы динамических ОЗУ SDRAM типа DDR2 (Double Data Rate Two Synchronous Dynamic Random Access Memory), которые образуют распределенную память базового модуля 16V5-75. Для организации распределенной памяти используется 20 микросхем MT47H32M16BN фирмы Micron Technology Inc. Каждая микросхема имеет структуру 2^{25} 16-разрядных слов. Общий объем распределенной памяти составляет 1,25 Гбайт. Кроме этого на базовом модуле установлено еще 128 Мбайт памяти, подключенной к контроллеру базового модуля – КБМ.

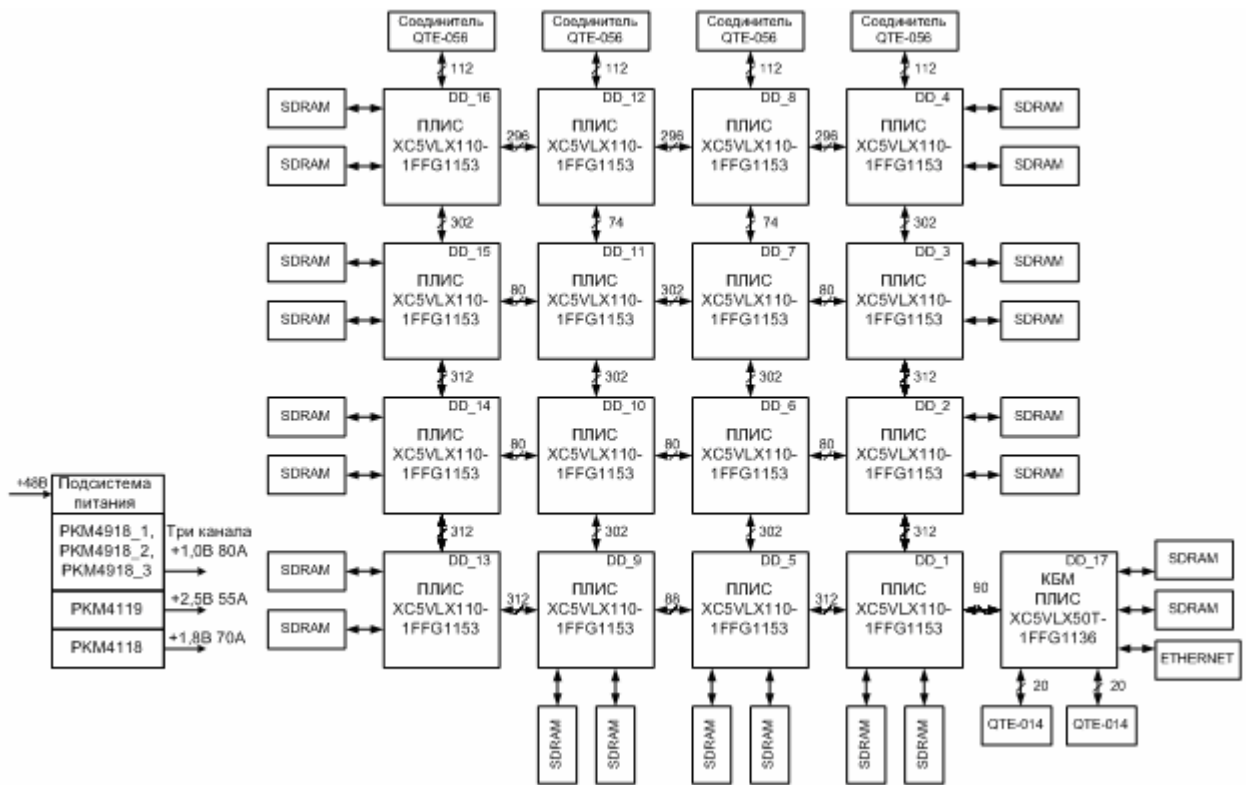


Рис. 6. Структура базового модуля 16V5-75

Распределенная память предназначена для хранения входных данных, промежуточных и конечных результатов обработки информации, а также фрагментов параллельных программ.

Основным назначением KBM являются функции управления подсистемами базового модуля, а также передача информации между базовым модулем и управляющим контроллером (ЭВМ типа IBM PC). Помимо этого KBM выполняет загрузку конфигурации ПЛИС решающего поля, участвует в диагностике устройств базового модуля, загружает входные данные в распределенную память базового модуля, выгружает результаты обработки, настраивает подсистему синхронизации, собирает и передает информацию о текущем состоянии базового модуля в управляющий контроллер. KBM выполнен на ПЛИС Virtex 5 XC5VLX50T-1FFG1136 фирмы Xilinx. Настройка конфигурации этой микросхемы осуществляется по включению питания из энергонезависимой памяти, в которой хранится конфигурационный файл KBM. Связь KBM с управляющим контроллером осуществляется посредством LVDS-каналов через два соединителя QTE-014, а также с помощью канала Ethernet.

Функция контроля температуры ПЛИС решающего поля осуществляется KBM путем опроса аппаратных блоков System Monitor, измеряющих температуру кристаллов, которые встроены в ПЛИС семейства Virtex 5.

Подсистема электропитания расположена непосредственно на базовом модуле и содержит портативные преобразователи напряжения серии РКМ фирмы Ericsson, на которые подается напряжение +48В от источника первичного питания. Для отвода тепла и поддержания

необходимых температурных режимов в микросхемах базового модуля предусмотрена комбинированная система охлаждения, которая включает радиаторы на ПЛИС решающего поля и вентиляторы для их обдува.

Таким образом, базовый модуль 16V5-75 представляет собой мощный вычислительный узел производительностью свыше 75 Гфлопс. На основе БМ 16V5-75 могут строиться вычислительные блоки, содержащие от одного до восьми базовых модулей производительностью от 75 до 600 Гфлопс. В то же время базовый модуль обладает достаточной автономностью и может легко комплексоваться с персональным компьютером типа IBM PC в качестве ускорителя и использоваться при решении различных задач.

Структура PBC-0.2-PC и PBC-0.2-VB

На основе базового модуля 16V5-75 в рамках проекта разработаны рабочая станция PBC-0.2-PC и вычислительный блок PBC-0.2-VB производительностью 300 Гфлопс. Основу этих изделий составляет многопроцессорный вычислительный блок (МВБ), который включает четыре базовых модуля, соединенных между собой в единый вычислительный ресурс быстрыми каналами LVDS и посредством коммутатора Ethernet.

Рабочая станция PBC-0.2-PC является представителем семейства PBC и предназначена для решения прикладных задач проектирования изделий микроэлектроники, управления в реальном времени сложными объектами, моделирования сложных технических и природных объектов и процессов, построения систем мониторинга, дистанционного зондирования и томографии и др.

Вычислительный блок PBC-0.2-VB практически полностью повторяет архитектуру рабочей станции PBC-0.2-PC, однако конструкции этих изделий значительно отличаются. Конструктивные отличия определяются назначением этих изделий: рабочая станция – это настольный вариант вычислительной системы, предназначенный для автономного использования, а вычислительный блок – это встраиваемый вариант, предназначенный для комплектования стоек СТ-1Р в составе представителей семейства PBC-1Р и PBC-5 и для создания суперЭВМ различных конфигураций. Вычислительный блок PBC-0.2-VB обладает теми же техническими параметрами, что и рабочая станция PBC-0.2-PC, и предназначен для решения перечисленных выше задач в составе PBC-1Р и PBC-5.

Реальная производительность PBC-0.2-PC и PBC-0.2-VB на задачах линейной алгебры, математической физики, цифровой обработки сигналов, символьной обработки составляет не менее 60% от пиковой производительности. Аппаратно-программные средства PBC-0.2-PC и PBC-0.2-VB позволяют динамически перестраивать архитектуру в процессе решения задачи на

двух уровнях: программном (на уровне элементарных процессоров и каналов распределенной памяти), обеспечивающем высокую скорость реконфигурации системы на задачи из данного класса; схемотехническом (на уровне логических ячеек ПЛИС), обеспечивающем модернизацию системы команд элементарных процессоров и высокую удельную производительность системы при переходе на задачи различных классов.

РВС-0.2-РС и РВС-0.2-ВБ обладают высокой надежностью и готовностью к работе.

Архитектурные отличия РВС-0.2-ВБ от РВС-0.2-РС заключаются в особенностях соединения базовых модулей в составе МВБ. На рис. 7а и 7б показана укрупненная структура рабочей станции РВС-0.2-РС и МВБ вычислительного блока РВС-0.2-ВБ. В МВБ рабочей станции базовые модули соединяются в кольцо, а в МВБ вычислительного блока базовые модули БМ0 и БМ3 имеют выходы за пределы РВС-0.2-ВБ с целью комплексирования нескольких вычислительных блоков РВС-0.2-ВБ в единую вычислительную структуру в составе стойки СТ-1Р. Во всем остальном архитектура РВС-0.2-ВБ совпадает с архитектурой РВС-0.2-РС.

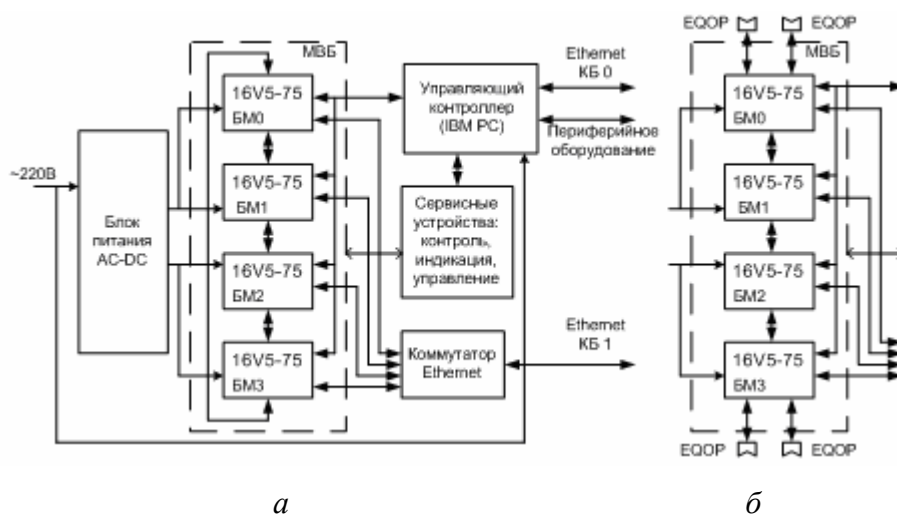


Рис. 7. Структура рабочей станции РВС-0.2-РС (а)
и МВБ вычислительного блока РВС-0.2-ВБ (б)

Управляющий контроллер, входящий в состав обоих изделий, представляет собой персональный компьютер типа IBM PC, встроенный в корпус РВС-0.2-РС или РВС-0.2-ВБ и функционирующий под управлением операционной системы Windows XP. Управляющий контроллер предназначен для: управления всеми компонентами РВС-0.2-РС или РВС-0.2-ВБ; хранения информации (исходной, промежуточной и результатов); трансляции параллельных программ с языков высокого уровня в исполняемые коды; тестирования и диагностики; подключения периферийного оборудования; подключения к локальным и глобальным информационным сетям посредством Ethernet.

Системы охлаждения PBC-0.2-PC и PBC-0.2-ВБ имеют некоторые отличия, связанные с особенностями их назначения и эксплуатации. Автономная работа рабочей станции позволяет использовать для охлаждения ПЛИС базовых модулей медные штыревые радиаторы и проточно-вытяжную вентиляцию корпуса PBC-0.2-PC, что, в свою очередь, приводит к уменьшению габаритов рабочей станции по сравнению с PBC-0.2-ВБ, а также к уменьшению потребляемой мощности и шумности. Вычислительный блок PBC-0.2-ВБ работает в более жестких условиях по сравнению с рабочей станцией из-за наличия фонового перегрева, создаваемого другими блоками и необходимостью прогонять воздушный поток не только через корпус блока, но и через корпус стойки, где он встречает дополнительное сопротивление. Поэтому для PBC-0.2-ВБ используется система с проточной вентиляцией корпуса блока, дополненная вентиляторами прямого обдува, установленными непосредственно на медных штыревых радиаторах ПЛИС базовых модулей.

Реконфигурируемая вычислительная система PBC-1P

Реконфигурируемая вычислительная система PBC-1P предназначена для: оснащения научных центров с целью проведения исследований в области физики, химии, биологии, космоса, построения информационно-управляющих систем для управления потенциально опасными производствами, решения задач аэрокосмической, автомобильной промышленности и энергетики. Пиковая производительность PBC-1P составляет 1200 Гфлопс.

Основу PBC-1P составляет вычислительная стойка СТ-1P, в которую монтируются четыре вычислительных блока PBC-0.2-ВБ, ПЭВМ, коммутатор Ethernet, система питания и система охлаждения. Компоновка стойки СТ-1P и PBC-1P представлена на рис. 1. Структурная схема PBC-1P показана на рис. 8.

Четыре вычислительных блока PBC-0.2-ВБ объединяются с помощью быстрых LVDS-каналов в единый вычислительный ресурс, содержащий до 16-ти базовых модулей 16V5-75 с общей пиковой производительностью 1,2 Тфлопс. Межблочные связи являются продолжением межмодульных и, в свою очередь, продолжением связей между ПЛИС решающих полей базовых модулей. В целом подобная организация логических связей реализует в составе стойки СТ-1P глобальный LVDS-канал передачи данных с единым темпом продвижения информации.

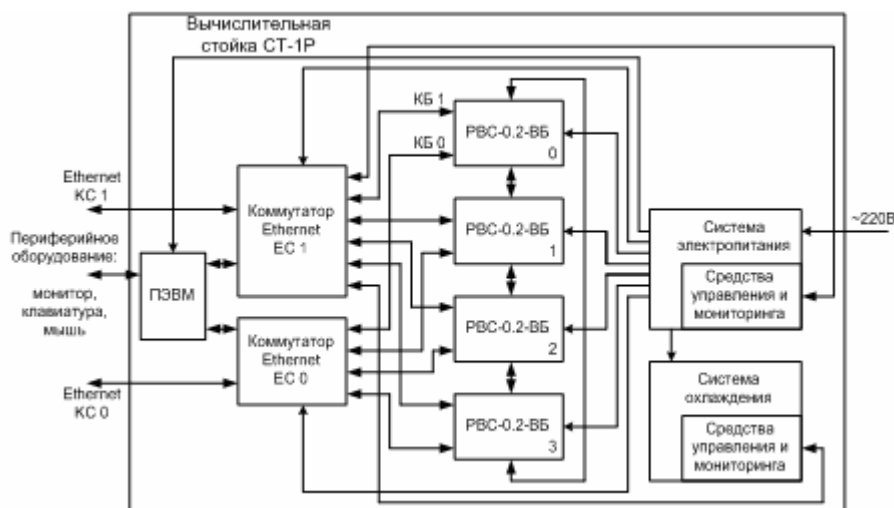


Рис. 8. Структурная схема РВС-1Р

Как было показано выше, каждый вычислительный блок РВС-0.2-ВБ имеет два канала Ethernet для связи с внешними сетями – КБ0 и КБ1. Для осуществления функций управления и мониторинга управляющие контроллеры вычислительных блоков по выходам КБ1 соединяются между собой посредством коммутатора ЕС1 под общим управлением ПЭВМ. Сетевые выходы КБ0 блока РВС-0.2-ВБ объединяются сетевым коммутатором стойки ЕС0, посредством которого можно установить прямые связи с любым из базовых модулей в составе стойки, минуя управляющие контроллеры блоков.

Такое соединение вычислительных ресурсов позволит максимально эффективно использовать возможности реконфигурируемой элементной базы вычислительной системы РВС-1Р. С одной стороны, быстрые каналы LVDS связывают решающие поля всех шестнадцати базовых модулей в единую вычислительную среду, позволяющую создавать многопроцессорную вычислительную систему со структурно-процедурной организацией вычислений в пределах четырех блоков РВС-0.2-ВБ, что дает возможность использовать все преимущества ресурсонезависимого программного обеспечения РВС. С другой стороны, система связей вычислительных блоков РВС-0.2-ВБ, благодаря сетевым технологиям и посредством коммутаторов Ethernet, позволяет вычислительной системе РВС-1Р приобретать черты кластерной ЭВМ, где в качестве кластеров могут выступать как блоки РВС-0.2-ВБ, так и базовые модули 16V5-75.

Встраиваемая ПЭВМ представляет собой промышленный компьютер типа IBM PC в корпусе, который приспособлен для включения в стойку СТ-1Р. ПЭВМ предназначена для управления работой четырех вычислительных блоков РВС-0.2-ВБ, управления работой коммутаторов Ethernet, управления работой системы питания и системы охлаждения, подключения периферийного оборудования: дисплея, клавиатуры, ручного манипулятора («мышь») и т.п., для подключения реконфигурируемой вычислительной системы РВС-1Р к

локальным и глобальным информационным сетям.

Реконфигурируемая вычислительная система РВС-5

Старшим представителем семейства РВС является изделие РВС-5 с пиковой производительностью 6000 Гфлопс.

Реконфигурируемая вычислительная система РВС-5 предназначена для использования в научно-исследовательских центрах с целью проведения исследований в области молекулярного моделирования, создания принципиально новых лекарственных препаратов и материалов нового поколения, построения информационно-моделирующих систем для долгосрочного прогнозирования погоды, моделирования задач сейсмоактивности, томографических исследований приповерхностных слоев Земли акустическими и электромагнитными волнами.

Упрощенная структурная схема реконфигурируемой вычислительной системы РВС-5 показана на рис. 9.

РВС-5 содержит пять стоек СТ-1Р, управляющую ЭВМ (УЭВМ), систему питания, систему охлаждения. К управляющей ЭВМ подключаются монитор, клавиатура, ручной манипулятор («мышь») и другие периферийные устройства.

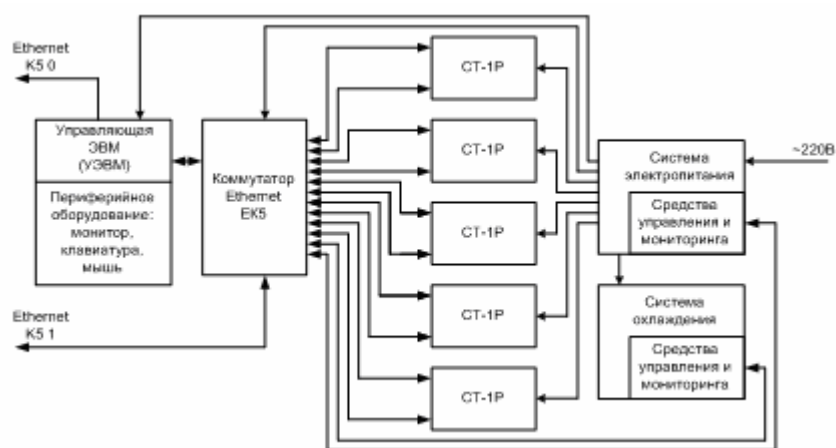


Рис. 9. Упрощенная структурная схема РВС-5

Основным вычислительным ресурсом РВС-5 являются пять стоек СТ-1Р, которые соединяются между собой посредством коммутаторов Ethernet с использованием сетевых технологий под общим управлением УЭВМ. Такое построение аппаратных средств позволяет, с одной стороны, связать воедино все вычислительные ресурсы стоек СТ-1Р, а с другой - максимально унифицировать оборудование, поскольку вычислительная часть РВС-5 строится из типовых узлов, уже использовавшихся для построения других представителей семейства РВС – базовых модулей 16V5-75, вычислительных блоков РВС-0.2-ВБ, стоек СТ-1Р.

Один из возможных вариантов соединения вычислительных ресурсов РВС-5 показан на рис. 10. В этом случае коммутатор Ethernet ЕС0 в составе стоек СТ-1Р подключается так, чтобы создать «горизонтальные» связи между блоками РВС-0.2-ВБ различных стоек системы РВС-5.

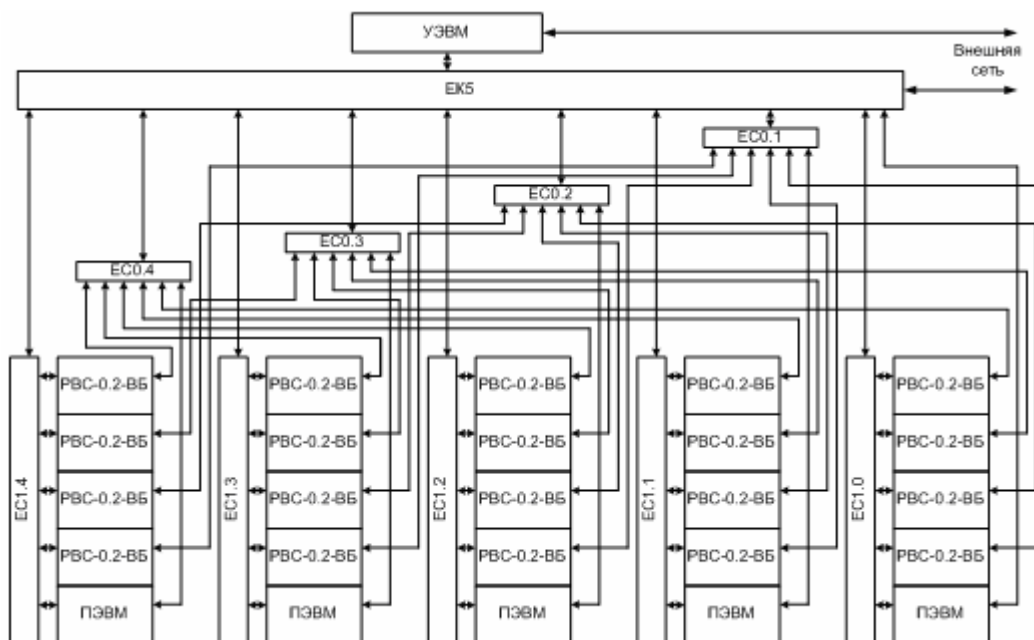


Рис. 10. Сетевое соединение вычислительных ресурсов РВС-5

Управляющая ЭВМ представляет собой персональный компьютер типа IBM PC с типовым набором необходимых устройств и периферийного оборудования и предназначена для управления работой вычислительных ресурсов пяти стоек СТ-1Р, управления работой системы питания и системы охлаждения, подключения периферийных устройств, а также для подключения РВС-5 к локальным и глобальным информационным сетям.

Реконфигурируемые ускорители персонального компьютера РУПК-50 и РУПК-25

Основным назначением реконфигурируемых ускорителей персонального компьютера РУПК-25 и РУПК-50 является наращивание возможностей персональных компьютеров типа IBM PC при решении вычислительно трудоемких задач, проектировании изделий микроэлектроники, математическом моделировании сложных технических и природных объектов и процессов, символьной обработке информации, оптимизации при эксплуатации нефтяных месторождений, дистанционном зондировании и томографии и др.

РУПК-50 и РУПК-25 предоставляют пользователю дополнительный вычислительный ресурс с пиковой производительностью свыше 50 и 25 Гфлопс соответственно. Реальная производительность РУПК-50 и РУПК-25 на задачах линейной алгебры, математической физики, цифровой обработки сигналов, символьной обработки превышает 60% от пиковой производительности. Аппаратно-программные средства позволяют динамически перестраивать

архитектуру РУПК-25 и РУПК-50 в процессе решения задачи на программном и схемотехническом уровнях.

РУПК-50 и РУПК-25 имеют одинаковую структуру, содержат одинаковые узлы и различаются только типом базового модуля (рис. 11).



Рис. 11. Структурная схема РУПК-50 (РУПК-25)

РУПК-50 (РУПК-25) содержит: базовый модуль 16V5-50 (16S3-25) с устройством сопряжения с персональным компьютером; блок питания AC-DC; панель индикации; систему охлаждения.

Базовые модули с динамически перестраиваемой архитектурой на основе ПЛИС для реконфигурируемых ускорителей персональных компьютеров строятся на тех же принципах, что и базовый модуль 16V5-75 старших представителей семейства PBC-0.2-PC, PBC-1P и PBC-5. Основным отличием от базового модуля 16V5-75 является более низкая стоимость и меньшая потребляемая мощность при сохранении высокой производительности. Поэтому базовые модули для РУПК-50 и РУПК-25 строятся на основе более дешевых электронных компонентов.

Решающее поле базового модуля 16V5-50 выполнено на 16-ти ПЛИС Virtex 5 XC5VLX110-1FFG1153 фирмы Xilinx, контроллер базового модуля – на ПЛИС XC5VLX50T-1FFG1136. Распределенная память реализована на 24-х микросхемах SDRAM типа DDR2 и имеет общий объем 1,5 Гбайта. В отличие от базового модуля 16V5-75 базовый модуль 16V5-50 не содержит каналов связи с другими базовыми модулями. Производительность базового модуля 16V5-50 составляет 50 Гфлопс.

Решающее поле базового модуля 16S3-25 выполнено на 16-ти ПЛИС Spartan 3 XC3S4000-5FG1156 фирмы Xilinx, каждая из которых содержит около $4 \cdot 10^6$ эквивалентных вентилях. Контроллер базового модуля выполнен на ПЛИС Virtex 4 XC4VFX60-10FF1152C. Распределенная память реализована на 24-х микросхемах SDRAM типа DDR2 и имеет общий объем 1,5 Гбайта. Производительность базового модуля 16S3-25 составляет 25 Гфлопс.

Обмен информацией РУПК-50 и РУПК-25 с персональным компьютером по данным и управлению осуществляется с помощью ПЛИС КБМ, используя LVDS- или Ethernet-каналы.

Система охлаждения поддерживает необходимый температурный режим электронных компонентов РУПК-50 (РУПК-25) и включает радиаторы и вентиляторы на температурно-напряженных микросхемах базового модуля, а также вентиляторы для принудительной прокачки воздуха в пределах всего объема корпуса.

Системное программное обеспечение

Программирование РВС отличается от программирования МВС традиционной архитектуры и его можно условно разделить на две составляющие: программирование структурное, которое создает необходимые вычислительные структуры в поле логических ячеек ПЛИС, и программирование процедурное – программирование в традиционном смысле, заключающиеся в организации вычислительного процесса в РВС. При этом программирование вычислительных структур вызывает у пользователей наибольшие трудности [10]. Это связано с тем, что пользователи традиционно привыкли программировать только организацию вычислительного процесса, опираясь на неизменяемую аппаратную поддержку средств вычислительной техники, в то время как для программирования вычислительных структур РВС требуется совершенно другая квалификация, а именно - квалификация схемотехника.

При программировании пользовательской задачи структура РВС приобретает черты специализированной многопроцессорной ЭВМ, которая оптимально соответствует структуре решаемой задачи из предметной области. Это обеспечивает высокую реальную производительность системы, близкую к пиковой производительности на широком классе задач, и позволяет достичь практически линейного роста производительности при наращивании аппаратного ресурса. Эффективность вычислительного процесса при реконфигурации архитектуры РВС на низком, схемотехническом уровне может быть повышена от 10 до 100 раз по сравнению с вычислительными системами, архитектура которых не может быть изменена. Это делает, с одной стороны, чрезвычайно привлекательными реконфигурируемые на низком уровне системы, а, с другой стороны, их программирование становится по сложности сопоставимым с созданием новой вычислительной системы. Такой подход требует новых методов и средств организации параллельных вычислительных процессов.

Создаваемое в рамках проекта системное и прикладное программное обеспечение обеспечивает потенциальным пользователям удобство программирования сложных практических задач на семействе РВС и включает: программный комплекс средств разработки прикладных программ, средства администрирования вычислительных ресурсов семейства РВС и служебные программы и драйверы.

Целью создания программного комплекса средств разработки прикладных программ семейства PBC является предоставление пользователю возможностей, которые позволяют создавать программы без привлечения специальных знаний в области схемотехники ПЛИС и по сложности будут приближены к обычному программированию для микропроцессоров и многопроцессорных ЭВМ. Основными задачами программного комплекса средств разработки является эффективная реализация вычислительно трудоемких фрагментов задач различных проблемных областей на произвольном количестве взаимосвязанных кристаллов ПЛИС и произвольном количестве базовых модулей, а также поддержка разработки и отладки прикладных программ на языках структурно-процедурных вычислений, в том числе с вызовом библиотечных функций настройки архитектуры системы и реализации необходимых вычислительных структур на множестве ПЛИС.

Программный комплекс средств разработки прикладных программ семейства PBC содержит: язык ассемблера Argus v.3.0; язык структурно-процедурного программирования высокого уровня COLAMO v.2.0; интегрированную среду разработки аппаратно-программных решений прикладных задач Argus IDE v.3.0, единую для всех представителей семейства PBC и поддерживающую языки программирования Argus v.3.0 и COLAMO v.2.0; отладчик параллельных программ на базовых модулях семейства PBC, поддерживающий межмодульные связи; программный интерфейс доступа к вычислительным ресурсам всех представителей PBC из различных сред программирования; среду разработки вычислительных структур для синтеза масштабируемых параллельно-конвейерных процедур, оперирующую библиотекой схемных решений (IP-ядер).

Язык структурно-процедурного программирования высокого уровня COLAMO [4, 7] обеспечивает синтаксическую поддержку реконфигурации аппаратной платформы PBC и возможность использования элементов библиотеки масштабируемых IP-ядер. Транслятор COLAMO v.2.0 для всех представителей семейства PBC обеспечивает трансляцию исходного кода программы в язык ассемблера Argus v.3.0 и в VHDL посредством среды разработки масштабируемых параллельно-конвейерных процедур Fire!Konstruktor, создавая тем самым конфигурационные файлы для ПЛИС.

Средства администрирования вычислительных ресурсов PBC, необходимые для обеспечения доступа и управления вычислительными ресурсами базовых модулей всех представителей семейства PBC, состоят из сервера удаленного доступа к вычислительным ресурсам, обеспечивающего обработку удаленных заявок на использование вычислительных ресурсов и поддерживающего очередь заявок с учетом приоритетов пользователей и клиентской части удаленного доступа, обеспечивающей прозрачный с точки зрения пользователя режим доступа и функционирования удаленных базовых модулей системы.

Одновременно с разработкой семейства РВС создается и будет развиваться инфраструктура для обучения потенциальных пользователей и организации доступа к библиотекам масштабируемых IP-ядер для различных предметных областей с целью их широкого продвижения на рынке высокопроизводительной вычислительной техники. За счет использования возможностей сервисов глобальной компьютерной сети Internet планируется обеспечить широкий доступ потенциальных пользователей к информационным ресурсам и программным продуктам, созданным в результате выполнения проекта.

Заключение

Описанные в статье принципы построения высокопроизводительных систем на основе реконфигурируемой элементной базы и создаваемое на этих принципах семейство РВС являются принципиально новым направлением развития высокопроизводительной вычислительной техники. В качестве основного вычислительного элемента в РВС используются не универсальные микропроцессоры, а ПЛИС. Это дает возможность пользователям создавать в базовой архитектуре РВС виртуальные специализированные вычислители, структура которых адекватна структуре решаемой задачи, что, в свою очередь, обеспечивает высокую эффективность вычислений и близкий к линейному рост производительности при наращивании вычислительного ресурса. Семейство РВС с динамически перестраиваемой архитектурой на основе ПЛИС предназначено для решения вычислительно трудоемких задач, решение которых на МВС традиционной архитектуры либо требует недопустимо больших временных затрат, либо вообще невозможно.

Проблема снижения производительности традиционных МВС на сильносвязанных задачах может быть разрешена только путем предоставления пользователю больших возможностей в части архитектурного программирования аппаратного параллелизма вычислительной системы. В этом контексте семейство РВС, описанию которого посвящена данная статья, обладает рядом преимуществ перед многопроцессорными системами традиционной организации, основанными на использовании микропроцессоров в качестве основного вычислительного элемента.

Литература

- 1 Аладышев О.С., Дикарев Н.И., Овсянников А.П. и др. СуперЭВМ: области применения и требования к производительности // Известия ВУЗов. Электроника, 2004. - №1. – С.13-17.
- 2 Воеводин В.В., Воеводин Вл.В. Параллельные вычисления. - С.-Петербург: «БХВ-Петербург», 2002. - 599 с.
3. Каляев А.В. Многопроцессорные вычислительные системы с программируемой архитектурой. - М.: Радио и Связь, 1984. - 240 с.

4. Каляев А.В., Левин И.И. Модульно-наращиваемые многопроцессорные системы со структурно-процедурной организацией вычислений. - М.: Янус-К, 2003. – 380 с.
5. Каляев И.А., Левин И.И. Многопроцессорные вычислительные системы (суперкомпьютеры): состояние и перспективы. // Вестник компьютерных и информационных технологий. – М.: Машиностроение, 2004. – №5. - С. 25-44.
6. Беседин И.В., Дмитренко Н.Н., Каляев И.А., Левин И.И., Семерников Е.А. Семейство базовых модулей для построения реконфигурируемых многопроцессорных вычислительных систем со структурно-процедурной организацией вычислений // Материалы Всероссийской научной конференции «Научный сервис в сети Интернет: технологии распределенных вычислений», г. Новороссийск. – М.: Издательство Московского университета, 2006. – С. 47-49.
7. Каляев И.А., Левин И.И. Высокопроизводительные модульно-наращиваемые многопроцессорные системы на основе реконфигурируемой элементной базы // Вычислительные методы и программирование. – М.: Изд-во Московского Университета, 2007. - Т.8. - №1. – С.181-190.
8. Каляев А.В., Каляев И.А., Левин И.И. Многопроцессорные вычислительные системы с программируемой архитектурой на основе ПЛИС // Вестник ЮНЦ РАН, 2004. – С.24-33.
9. Дордопуло А.И., Каляев И.А., Левин И.И., Семерников Е.А. Семейство многопроцессорных вычислительных систем с динамически перестраиваемой архитектурой // Материалы Четвертой Международной научной молодежной школы «Высокопроизводительные вычислительные системы». - Таганрог: Изд-во ТТИ ЮФУ, 2007. – С. 68-74.
10. Левин И.И. Язык параллельного программирования высокого уровня для структурно-процедурной организации вычислений // Труды Всероссийской научной конференции. - М.: Изд-во МГУ, 2000. – С.108-112.